ASYMPTOTIC ANALYSIS OF RANDOM MATRICES WITH EXTERNAL SOURCE AND A FAMILY OF ALGEBRAIC CURVES

K. T.-R. MCLAUGHLIN

ABSTRACT. We present a set of conditions which, if satisfied, provide for a complete asymptotic analysis of random matrices with source term containing two distinct eigenvalues. These conditions are shown to be equivalent to the existence of a particular algebraic curve. For the case of a quartic external field, the curve in question is proven to exist, yielding precise asymptotic information about the limiting mean density of eigenvalues, as well as bulk and edge universality.

1. INTRODUCTION

1.1. **Random Matrices with Source.** Consider the probability measure

$$(1) \qquad \mu_n(dM) = \frac{1}{Z_n} e^{-n \, \text{Tr}(V(M)-AM)} dM,$$

defined on $n \times n$ Hermitian matrices, M. Here dM denotes Lebesgue measure on the matrix entries, $dM = \prod_{j=1}^n dM_{jj} \prod_{1 \leq j < k \leq n} d\text{Re}(M_{jk})d\text{Im}(M_{jk})$, the matrix A is a fixed $n \times n$ matrix, and the parameter Z_n is a normalization constant chosen so that (1) is a probability measure.

This defines an ensemble of random matrices, in which the matrix A plays the role of an external source. The function V should be real and grow sufficiently rapidly that the above measure possesses all finite moments; it is typically assumed to be a polynomial.

The family of measures described by (1) are not invariant under unitary transformations, and the analysis of eigenvalue statistics under these measures is in its infancy in comparison to the so-called "unitary ensembles" for which a great deal is known.

Following the work of Brézin-Hikami [5], [6], and P. Zinn-Justin [13] [14], a team of researchers (Bleher and Kuijlaars [2], [3],[4] and Aptekarev, Bleher and Kuijlaars [1]) have considered the large n behavior of eigenvalue statistics under (1) from the point of view of Riemann–Hilbert problems.

In [13], Zinn-Justin showed that the eigenvalues of such random matrices are in fact a determinantal point process, and in [2], the authors showed that this representation may be described in terms of the solution of a matrix Riemann–Hilbert problem, the size of which depends on the number of distinct eigenvalues of the matrix A.

In the present paper we shall assume that the matrix A possesses two eigenvalues, $-a$ and a, with multiplicities n_1 and n_2, respectively (with $n_1 + n_2 = n$). In this case, the Riemann–Hilbert problem is as follows:

Riemann-Hilbert Problem 1.1. (a) *A is analytic on $\mathbb{C} \setminus \mathbb{R}$.*

(b) *The boundary values of A satisfy*

$$(2) \qquad A_+(x) = A_-(x) \begin{pmatrix} 1 & w_1 & w_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad for \ z \in \mathbb{R}$$

where

$$(3) \qquad w_1 = e^{-nV_1} = e^{-n(V(x)+ax)}, \quad w_2 = e^{-nV_2} = e^{-n(V(x)-ax)}.$$

(c) *As $z \to \infty$, we have*

(4)
$$A(z) = \left(I + O\left(\frac{1}{z}\right)\right) \begin{pmatrix} z^n & 0 & 0 \\ 0 & z^{-n_1} & 0 \\ 0 & 0 & z^{-n_2} \end{pmatrix}.$$

In terms of this Riemann–Hilbert problem, the probability measure on eigenvalues induced by (1) may be re-written as follows:

(5)
$$d\mu_n(x_1, \ldots, x_n) \quad = \quad \det \ (K_n(x_i, x_j))_{1 \leq i,j \leq n} \quad d^n x,$$

[13], [2] where now

$$K_n(x, y) \quad = \quad \frac{e^{-n(V(x)+V(y))}}{2\pi i(x-y)} \left\{ e^{nay} \left[Y(y)^{-1} Y(x) \right]_{21} \right.$$
$$\left. + e^{-nay} \left[Y(y)^{-1} Y(x) \right]_{31} \right\}$$

The formula (5) is not just a concise representation of the probability measure, it turns out that all statistical properties of eigenvalues can be related to K_n. For example,

$$\text{Prob } \{\text{no eigenvalues in } (a, b)\} \quad = \quad \det \ (1 - \mathcal{K}_n)_{L^2[(a,b)]}$$

where $\mathcal{K}_n$ is the integral operator with kernel $K_n(x, y)$:

$$\mathcal{K}_n f \quad = \quad \int_a^b K_n(x, y) f(y) dy.$$

The representation of the correlation functions in terms of the kernel K_n is due to Zinn-Justin [13], and the representation of this kernel in terms of a Riemann–Hilbert problem is due to Bleher and Kuijlaars [2].

Remark: In the case that the matrix A possesses p distinct eigenvalues with multiplicities $n_1, n_2, \ldots, n_p$, the above representations generalize directly, but the associated Riemann–Hilbert problem is $p + 1 \times p + 1$.

In [5] [6], the authors considered the case that V is a quadratic. In the series of papers [3], [1], and [4], the authors also considered the Gaussian case, but from the point of view of Riemann–Hilbert problems. The goal of these works was to study the behavior of the eigenvalue statistics in the large n limit. In [14], Zinn-Justin also studied the large n limit, for the case of general V. In that work, one important issue was the description of the limiting density of states in terms of a function for which an existence theorem is lacking. He explains very carefully the analyticity properties and branch-cut structure required of this function, in the so-called "one-cut" case, under the assumption that the density of the eigenvalues of A has a smooth limit when n tends to ∞.

Here we consider the case that A possesses two eigenvalues, of multiplicities n_1 and n_2, and our interest is in the behavior when $n, n_1, n_2 \to \infty$ so that $\frac{n_j}{n} \to x_j$. The main goal of our work is to provide a new and explicit characterization of the limiting density of states in terms of an algebraic curve. The following is a summary of the results in this paper:

(1) In Section 2, we present a list of conditions (referred to as an "ideal situation") which, if true, yield a transformation of the Riemann–Hilbert problem (1.1) to a "normal form", that is, a Riemann–Hilbert problem from which subsequent asymptotic analysis (for $n \to \infty$) is straightforward.

(2) In Section 3, we show that if these conditions are satisfied, then there exists an algebraic curve whose roots yield the desired transformation. The curve is always of the form

(6)
$$w^3 - V'(z)w^2 + \mathcal{C}_1(z)w - \mathcal{C}_0(z) = 0,$$

where $\mathcal{C}_1$ and $\mathcal{C}_2$ are analytic functions of z.

(3) For arbitrary polynomial external fields, the algebraic curve may be determined up to a finite number of free parameters (see Section 4).

(4) In Section 6, we consider the Gaussian and Quartic cases. For the Gaussian case, this algebraic curve has no free parameters, and in the case $n_1 = n_2 = n/2$, is equivalent to the algebraic curve used by Bleher and Kuijlaars [3, 4] and by Aptekarev, Bleher and Kuijlaars [1].

(5) For the quartic case, $V = x^4/4$, with $n_1 = n_2 = n/2$, we prove that for a sufficiently large there exists a choice of the free parameters so that the algebraic curve yields the desired transformation. Each of the two relevant roots possesses a single branch cut, which is the "one-cut" case.

(6) In Section 7, in the "one-cut" case, we show that *if the desired transformation exists* bringing the Riemann–Hilbert problem into a normal form for subsequent Riemann–Hilbert analysis, then the analysis of Bleher and Kuijlaars [3] can be applied directly.

(7) It follows that if the desired transformation exists, then (i) the limiting mean density of eigenvalues exists and is described in terms of the algebraic curve, (ii) bulk universality holds true, and (iii) edge universality hold true (see Section 8).

(8) The combined results of Sections 6 and 7 show that for the quartic case $V(x) = x^4/4$, with n even and $n_1 = n_2 = n/2$, bulk and edge universality still hold true, for all a sufficiently large.

Remark: In the analysis of singular limits of 2×2 Riemann–Hilbert problems associated to integrable systems and random matrix theory, there has often arisen an existence problem; a transformation is required satisfying a number of nonlocal properties, in order to transform to a new Riemann–Hilbert problem from which subsequent asymptotic analysis is straightforward (see, for example, [7]). The existence of this transformation has been shown to be equivalent to the existence of an equation of the form

$$F^2 + \mathcal{W}(z)F + \mathcal{U}(z) = 0,$$

in which (i) the functions $\mathcal{W}$ and $\mathcal{U}$ are analytic and defined implicitly in terms of the desired transformation, (ii) all branch points lie on the contour of the Riemann–Hilbert problem, and (iii) the roots $F_{\pm}(z)$ of this equation satisfy a variety of inequalities and relations on various subsets of the contour.

The proofs of existence of these transformations (or equations) are nontrivial, and require techniques ranging from WKB analysis of odes, through the theory of level trajectories of quadratic differentials, to the theory of logarithmic potentials and equilibrium measures.

Viewed in this light, the present paper explains that for the analysis of 3×3 Riemann–Hilbert problems, the existence of a suitable transformation is equivalent to the existence of an equation of the form (6) whose roots also satisfy a host of properties. For quartic potentials, we establish an existence theorem via algebraic geometry.

2. LARGE n ASYMPTOTIC ANALYSIS OF THE RIEMANN-HILBERT PROBLEM 1.1

This paper studies the large n asymptotic behavior of the solution to the Riemann–Hilbert problem 1.1, under the assumption that the functions V_j are entire. Furthermore, we will assume that the integers n_j grow to ∞ with n, in such a way that

$$(7) \qquad \frac{n_j}{n} \to x_j, \quad j = 1, 2.$$

2.1. **First step in the asymptotic analysis.** In analogy with the more standard case of orthogonal polynomials (see, for example, [7]-[8]), we seek a vector of functions

$$(8) \qquad \mathbf{G}(z) = (g_0(z), g_1(z), g_2(z))$$

satisfying the following set of assumptions.

(1) There are 2 disjoint sets, I_1 and I_2, each a compact subset of $\mathbb{R}$, with δ_j denoting the supremum of each set.

(2) For each $j = 1, 2$, the function g_j is analytic in $\mathbb{C} \setminus (-\infty, \delta_j)$.

(3) For $j = 1, 2$, each function g_j behaves as follows at ∞:

$$(9) \qquad g_j(z) = \log z + \mathcal{O}\left(z^{-1}\right).$$

(4) The functions g_j satisfy

$$(10) \qquad g_0 = g_1 + g_2.$$

In addition to these basic conditions, there will be a collection of important conditions on these functions, collectively referred to as the "Ideal Situation". The goal is to deduce that there exists a unique vector $\mathbf{G}$ that can be used to transform the Riemann-Hilbert problem 1.1 to a new Riemann-Hilbert problem in a form suitable for asymptotic analysis. To that end, the transformation that we seek is as follows:

$$(11) \qquad B := \text{diag}\left(e^{-n\hat{\ell}_0/2}, e^{n_1\hat{\ell}_1/2}, e^{n_2\hat{\ell}_2/2}\right) A(z)\, \text{diag}\left(e^{-n(g_0-\hat{\ell}_0/2)}, e^{n_1(g_1-\hat{\ell}_1/2)}, e^{n_2(g_2-\hat{\ell}_2/2)}\right),$$

where $\hat{\ell} = (\hat{\ell}_0, -\hat{\ell}_1, -\hat{\ell}_2)$ is a vector of constants, which will be chosen along with the vector $\mathbf{G}$. For now, the only restriction on these constants is that $n\hat{\ell}_0 = n_1\hat{\ell}_1 + n_2\hat{\ell}_2$.

The matrix B inherits jump relationships across $\mathbb{R}$, that are determined by the Riemann-Hilbert problem 1.1 for A, along with the boundary value behavior of the vector $\mathbf{G}$. Therefore, B satisfies the following Riemann–Hilbert problem.

Riemann-Hilbert Problem 2.1.   (a) *B is analytic on $\mathbb{C} \setminus \mathbb{R}$.*

    (b) *The boundary values of B satisfy*

$$(12) \qquad B_+(x) = B_-(x) \begin{pmatrix} e^{-n(g_0^+ - g_0^-)} & e^{ng_0^- + n_1 g_1^+ - nV_1 - n_1\ell_1} & e^{ng_0^- + n_2 g_2^+ - nV_2 - n_2\ell_2} \\ 0 & e^{n_1(g_1^+ - g_1^-)} & 0 \\ 0 & 0 & e^{n_2(g_2^+ - g_2^-)} \end{pmatrix},$$

    (c) *As $z \to \infty$, we have*

$$(13) \qquad B(z) = \left(I + O\left(\frac{1}{z}\right)\right).$$

Note that for ease of notation, we have introduced the constants ℓ_j, $j = 1, 2$ (without hats). These are defined by $\ell_1 = \frac{n\hat{\ell}_0 + n_1\hat{\ell}_1}{2n_1}$, and $\ell_2 = \frac{n\hat{\ell}_0 + n_2\hat{\ell}_2}{2n_2}$.

Ideal Situation

In analogy with the case of orthogonal polynomials, a detailed asymptotic analysis would be possible if the following were true.

    A. (As mentioned before) There are 2 disjoint sets, I_1 and I_2, each a compact subset of $\mathbb{R}$, so that for each $j = 1, 2$, g_j is analytic on $\mathbb{C} \setminus (-\infty, \delta_j)$ (where δ_j is the supremum of the set I_j).

    B. The function g_1 and its derivative g_1' are continuous in the closure of $\mathbb{C} \setminus (-\infty, \delta_1)$, and the functions g_2 and g_2' are continuous in the closure of $\mathbb{C} \setminus (-\infty, \delta_2)$. That is, g_j and g_j' achieve their boundary values in the sense of continuous functions.

    C. For $j = 1, 2$, we have that for $z \in I_j$, the following two properties hold true: (i) $\frac{1}{i}\left(g_j^+ - g_j^-\right)$ is real, and decreasing as one traverses I_j according to its orientation, and (ii)

$$(14) \qquad ng_0^- + n_j g_j^+ - nV_j - n_j\ell_j = 0.$$

    D. For $j = 1, 2$, we have that for $z \in \mathbb{R} \setminus I_j$, the following two properties hold true: (i) $\frac{1}{i}\left(g_j^+ - g_j^-\right)$ is constant, and (ii)

$$(15) \qquad \text{Re}\left[ng_0^- + n_j g_j - nV_j - n_j\ell_j\right] < 0.$$

Remark: The reader may wonder why conditions A-D above have been described as an ideal situation. The explanation requires considerable calculation, but the basic idea is this: on each subinterval I_j of $\mathbb{R}$, the jump matrix may be factored in a manner quite analogous to the matrix factorization that has been used for asymptotic analysis of 2×2 Riemann-Hilbert problems in approximation theory over the past 8 years. The theory for the case of 2×2 matrices can be essentially embedded in the more general setting. The real evidence that this works is in the papers [9], [3], [1], and [4], where the authors prove in special cases that the conditions A-D can be achieved, and then proceed to use this information to compute a complete asymptotic description for the Riemann-Hilbert problem.

In Section 7, we will carry out this Riemann–Hilbert analysis, for the "one-cut" case. Indeed, the reader will see that in this case, the subsequent Riemann–Hilbert analysis is virtually identical to the analysis carried out by Bleher and Kuijlaars [3].

Remark: The "ideal situation" described in A-D above does not always hold true. Even in the Gaussian case, for $a < 1$, it turns out that the desired sets I_1 and I_2 are not subsets of $\mathbb{R}$, but rather they are piecewise analytic arcs (actually each is a union of two line segments). The more general setting can be described as follows

- The Riemann–Hilbert problem 1.1 is modified so that the jump is not across the real axis, but rather is across two contours, Γ_1 and Γ_2.
- The ideal situation would be if the sets I_j, no longer subintervals of $\mathbb{R}$, become unions of piecewise analytic arcs, subsets of the contours Γ_1 and Γ_2, so that the conditions described in A-D above are true, on the sets I_j. (The conditions in B must be generalized so that the functions g_j are analytic with continuously differentiable boundary values off semi-infinite subsets of the contours Γ_1 and Γ_2.)

For V convex and a sufficiently large, it seems to be the case that the more down-to-earth conditions of A-D, with most action taking place on the real axis, should capture the situation, but this remains to be seen. On the other hand, when the more general setting does occur, the existence of an algebraic curve may be deduced as well, and the actual form of the equation depends only on the external field V, and the parameters a, n_1 and n_2.

Using the fact that $ng_0 = n_1 g_1 + n_2 g_2$, we may rewrite the condition (14) as

$$(16) \qquad n_j \left(g_j^- + g_j^+ \right) + n_k g_k - nV_j - n_j \ell_j = 0, \quad z \in I_j, \quad k, j = 1, 2, \quad k \neq j.$$

In the next Section we will show that the collection of boundary relations contained in (16) are intimately related to an equation of the form

$$(17) \qquad w^3 + a_2(z)w^2 + a_1(z)w + a_0(z) = 0$$

in which the coefficients $a_j(z)$ are analytic for all $z \in \mathbb{C}$. We will show that the roots of this polynomial are expressible explicitly in terms of the g_j's and V_j's.

Remark: For the case of $p \times p$ Riemann–Hilbert problems associated to Random Matrices with Source, the collection of boundary relations analogous to (16) should be related to an equation of the form

$$(18) \qquad w^p + a_{p-1}(z)w^{p-1} + \cdots + a_0(z) = 0.$$

3. An algebraic curve

We begin by defining f_1 and f_2 as follows:

$$(19) \qquad f_j(z) = \frac{n_j}{n} g_j'(z), \quad j = 1, 2.$$

The boundary relations (16) yield, upon differentiation, that f_1 and f_2 satisfy

$$(20) \qquad \left(f_1^- + f_1^+ \right) + f_2 - V_1' = 0, \quad z \in I_1,$$
$$(21) \qquad \left(f_2^- + f_2^+ \right) + f_1 - V_2' = 0, \quad z \in I_2.$$

In addition, if the condition that $\frac{1}{i} \left(g_j^+ - g_j^- \right)$ is constant on $\mathbb{R} \setminus I_j$ holds true (see part D of the "ideal situation" of the previous section) then clearly for each $j = 1, 2$, f_j is analytic in $\mathbb{R} \setminus I_j$.

Lemma 3.1. *Assuming the hypotheses A - D set forth in the "ideal situation" of the previous section, we conclude that the functions f_j satisfy the following relations for all $z \in \mathbb{C}$.*

$$(22) \qquad f_1(z)^2 + f_1(z)f_2(z) + f_2(z)^2 - f_1(z)V_1'(z) - f_2(z)V_2'(z) = \mathcal{A}(z)$$
$$(23) \qquad f_1(z)f_2(z)\left(f_1(z) + f_2(z)\right) - V_1'(z)f_1(z)\left(V_1'(z) - f_1(z)\right) - V_2'(z)f_2(z)\left(V_2'(z) - f_2(z)\right) = \mathcal{B}(z),$$

where $\mathcal{A}(z)$ and $\mathcal{B}(z)$ are entire functions, implicitly defined as follows:

$$(24) \qquad \mathcal{A}(z) = \frac{1}{2\pi i} \int_\gamma \frac{f_1(s)V_1'(s) + f_2(s)V_2'(s)}{s - z} ds$$

$$(25) \qquad \mathcal{B}(z) = \frac{1}{2\pi i} \int_\gamma \frac{V_1'(s)f_1(s)\left(V_1'(s) - f_1(s)\right) + V_2'(s)f_2(s)\left(V_2'(s) - f_2(s)\right)}{s - z} ds$$

and in both (24) and (25), the contour of integration γ may be taken to consist of a finite union of simple closed Jordan curves which do not intersect and whose interiors are disjoint, each oriented in the clockwise direction, one subset of these contours encircling the intervals comprising I_1, a second subset encircling the intervals comprising I_2, and one final contour encircling the point z. (Of course, since V_1 and V_2 are entire,

this may be deformed into one contour, a large circle oriented in the clockwise direction, which encircles the sets I_1 and I_2, and also encircles the point z.)

Proof

Let $H_2 = f_1(z)^2 + f_1(z)f_2(z) + f_2(z)^2 - f_1(z)V_1'(z) - f_2(z)V_2'(z)$. As defined, H_2 is analytic in $\mathbb{C} \setminus (I_1 \cup I_2)$. However, the jump of H_2 across either of these sets may be computed. Indeed, across I_1, we have

$$(26) \qquad H_2^+(z) - H_2^-(z) = \left(f_1^+(z) - f_1^-(z) \right) \left\{ f_1^+(z) + f_1^-(z) + f_2(z) - V_1'(z) \right\} \quad = \quad 0.$$

where in the last equality, we have used the boundary relation (20). In exactly the same way one may verify that H_2 also possesses no jump across I_2. Since we know that f_1 and f_2 are bounded, H_2 can have no isolated singularities, and so Morera's theorem tells us that H_2 is in fact entire. Equation (22) with $\mathcal{A}(z)$ defined in (24) follows by Cauchy's theorem.

The proof of (23) goes along the same lines. One begins by defining

$$(27) \quad H_3 = f_1(z)f_2(z) \left(f_1(z) + f_2(z) \right) - V_1'(z)f_1(z) \left(V_1'(z) - f_1(z) \right) - V_2'(z)f_2(z) \left(V_2'(z) - f_2(z) \right),$$

and observing that H_3 also possesses no jumps across $I_1 \cup I_2$, and hence it must be entire. Then (23) with $\mathcal{B}(z)$ defined in (25) follows by Cauchy's theorem. $\square$

Remark: The reader will note that the assumption that V_1 and V_2 should be entire is for comfort; for example, if V_1' and V_2' are meromorphic functions, then the only modification to the theorem is that $\mathcal{A}(z)$ and $\mathcal{B}(z)$ are also meromorphic, as they might have poles at those points that are poles of V_1' and V_2'. The integral representations (24) and (25) for $\mathcal{A}$ and $\mathcal{B}$ remain true, with the contour of integration taken to consist of 3 simple closed Jordan curves that do not intersect each other, each oriented in the clockwise direction, one encircling I_1, the second encircling I_2, and the third encircling the point z. (It may not be possible to deform these into a single contour without introducing additional residues.)

Remark: In order to conclude that H_2 can have no isolated singularities on I_1 or I_2, we don't actually need that the boundary values be bounded. The Lemma remains true if the boundary values of H_2 lie in $L^p(I_j)$ for some $1 < p < 2$. This may be translated into conditions on the functions f_j, but this is, at the present time, not particularly useful.

Remark: We note in passing that not all of the assumptions of A-D are required for the proof of this Lemma, and in fact what is actually required is A, B, (14), and item (i) of D.

The following uniqueness result follows immediately from Lemma 3.1.

Theorem 3.1. *Suppose that there are two disjoint sets, I_1 and I_2, each one bounded and consisting of a disjoint union of intervals. Then the boundary value problem for (20)-(21) (with, for $j = 1, 2$, f_j analytic in $\mathbb{C} \setminus I_j$, and the boundary values of f_j being bounded) possesses at most one solution.*

Proof If there is another pair of functions $\hat{f}_1$ and $\hat{f}_2$ that satisfy the same boundary value relations and asymptotics, then their difference $F_j := f_j - \hat{f}_j$, $j = 1, 2$, satisfies the following relations.

$$(28) \qquad\qquad F_1^+ + F_1^- + F_2 = 0, \quad z \in I_1,$$

$$(29) \qquad\qquad F_2^+ + F_2^- + F_1 = 0, \quad z \in I_2,$$

$$(30) \qquad\qquad F_j = \frac{c_j}{z^2}, z \to \infty.$$

Thus, Lemma 3.1 implies that for all $z \in \mathbb{C}$,

$$(31) \qquad\qquad\qquad F_1^2 + F_1 F_2 + F_2^2 = 0,$$

$$(32) \qquad\qquad F_1 F_2 \left(F_1 + F_2 \right) = 0.$$

These identities finally imply that $F_j \equiv 0$ for $j = 1, 2$, and we have proven uniqueness. $\square$

Lemma 3.1 clearly demonstrates that the boundary value relations (20)-(21) are related to algebraic equations; it is useful to recast these relations in terms of the following result.

Theorem 3.2. *Define* $r_1(z)$, $r_2(z)$, *and* $r_3(z)$ *as follows:*

$$(33) \qquad r_1 = f_1 - \frac{1}{2}\left(V_1' - V_2'\right)$$

$$(34) \qquad r_2 = f_2 + \frac{1}{2}\left(V_1' - V_2'\right)$$

$$(35) \qquad r_3 = -(r_1 + r_2) + \frac{1}{2}\left(V_1' + V_2'\right).$$

Then, under the assumptions of Lemma 3.1, the following holds true:

$$(36) \qquad E(w, z) := (w - r_1)(w - r_2)(w - r_3) = w^3 - \mathcal{C}_2(z)w^2 + \mathcal{C}_1(z)w - \mathcal{C}_0(z),$$

where

$$(37) \qquad \mathcal{C}_2(z) = \frac{1}{2}(V_1'(z) + V_2'(z))$$

$$(38) \qquad \mathcal{C}_1(z) = \frac{-1}{4}\left(4\mathcal{A}(z) + (V_1'(z) - V_2'(z))^2\right)$$

$$(39) \qquad \mathcal{C}_0 = \frac{1}{8}\left[4\mathcal{A}(z)(V_1'(z) + V_2'(z)) - 8\mathcal{B}(z) - (V_1'(z) - V_2'(z))^2(V_1'(z) + V_2'(z))\right].$$

Proof

The proof is a straightforward exercise in algebra: one must compute the coefficients of each power of w in the expression for E, using the definition (33)-(35) of r_j, along with (22) and (23). So, for example, (37) may be verified as follows:

$$(40) \qquad \mathcal{C}_2(z) = r_1 + r_2 + r_3 = \frac{1}{2}(V_1'(z) + V_2'(z)),$$

from (35). The relations (38) and (39) may be verified in a similar fashion. $\square$

Remark: As with Lemma 3.1, the assumption that V_1 and V_2 should be entire is for convenience, and in particular if V_1' and V_2' are meromorphic, then the Theorem remains true (but with $\mathcal{A}$ and $\mathcal{B}$ adjusted to admit the possibility of poles at the pole locations of V_1' and V_2').

Remark: It is important to observe that Theorem 3.2 yields nontrivial information about the functions g_1 and g_2. Indeed, the following result provides a drastic reduction in the possible complexity of the sets I_1 and I_2.

Theorem 3.3. *Assuming the hypotheses A - D set forth in the "ideal situation" of the previous section, each of the sets I_1 and I_2 consists of at most a finite union of intervals.*

Proof Since f_1 is analytic in a neighborhood of ∞, and since I_1 is a contour with finite arclength, f_1 may be expressed as a Cauchy integral of its jumps across I_1. The same holds true for f_2, and so we have

$$(41) \qquad f_j(z) = \frac{1}{2\pi i}\int_{I_j}\frac{f_j^+(s) - f_j^-(s)}{s - z}ds, \quad j = 1, 2.$$

On the other hand, the functions r_1, r_2, and r_3 are the roots of a function $E(z, w)$ that is a cubic polynomial in w with coefficients that are entire functions of z, and so its roots are analytic functions of z, with at worst a finite number of branching points in any finite part of $\mathbb{C}$. From the definition of the r_j's, we learn that the functions f_1 and f_2 can have at most a finite number of branching points in the entire plane. Each endpoint of a sub-arc of I_j clearly corresponds to a branching point, and so we conclude that I_1 and I_2 must each consist of at most a finite union of intervals. $\square$

The conditions A through D above, when satisfied, should in principle allow one to carry out a complete asymptotic analysis of the associated multiple orthogonal polynomials, and obtain complete control of the asymptotic statistics of the eigenvalues. What has not been pinned down completely under these conditions is the behavior near the endpoints of I_1 and I_2. This edge behavior depends on the order of vanishing of the quantities $ng_0^- + n_j g_j^+ - nV_j - n_j\ell_j$ and $\frac{1}{i}\left(g_j^+ - g_j^-\right)$ at the endpoints. The following additional condition, expected to be the generic case, implies that all edge behaviors are governed by the Airy equation. We will take (α, β) to be one of the intervals comprising the set I_j (which is now taken to consist of a finite union of disjoint intervals).

E 1. The behavior near each left endpoint α is as follows. There is $c_\alpha > 0$ so that
- $g_0^-(z) + x_j g_j^+(z) - V_j(z) - x_j \ell_j = -c_\alpha |z - \alpha|^{3/2} + \mathcal{O}\left(|z - \alpha|^{5/2}\right)$ for z near α and $z < \alpha$.
- $\frac{1}{i}\left(g_j^+(z) - g_j^-(z)\right) = c_\alpha |z - \alpha|^{3/2} + \mathcal{O}\left(|z - \alpha|^{5/2}\right)$ for z near α and $z > \alpha$.

E 2. The behavior near each right endpoint β is as follows. There is $c_\beta > 0$ so that
- $g_0^-(z) + x_j g_j^+(z) - V_j(z) - x_j \ell_j = -c_\beta |z - \beta|^{3/2} + \mathcal{O}\left(|z - \beta|^{5/2}\right)$ for z near β and $z > \alpha$.
- $\frac{1}{i}\left(g_j^+(z) - g_j^-(z)\right) = c_\beta |z - \beta|^{3/2} + \mathcal{O}\left(|z - \beta|^{5/2}\right)$ for z near β and $z < \beta$.

4. Polynomial external fields

According to the notation and definitions of Sections 2 and 3, we have $V_1' = V'(x) + a$ and $V_2' = V'(x) - a$, and hence one may verify that the desired algebraic curve has the form

$$(42) \qquad w^3 - (V'(z))\, w^2 - \left[\mathcal{A} + a^2\right] w - \left[\left(\mathcal{A} - a^2\right) V'(z) - \mathcal{B}\right] = 0.$$

The quantities $\mathcal{A}$ and $\mathcal{B}$ are entire functions, as yet undetermined (aside from the fact that they satisfy (24) and (25)).

Now if it is the case that $V(x)$ is a polynomial, more information can be obtained about $\mathcal{A}$ and $\mathcal{B}$. In this case, they satisfy

$$(43) \qquad \mathcal{A}(z) = \mathrm{Poly}\left[-V'(z)\left(f_1(z) + f_2(z)\right)\right],$$

$$(44) \qquad \mathcal{B}(z) = \mathrm{Poly}\left[-\left(V'(z)\right)^2 \left(f_1(z) + f_2(z)\right) + V'(z)\left(f_1^2 + 2a\left(f_2 - f_1\right) + f_2^2\right)\right]$$

where $\mathrm{Poly}(Q)$ refers to the polynomial part of the Laurent expansion of Q valid for z near ∞. Clearly then, $\mathcal{A}$ is a polynomial of degree $\deg(V) - 2$, and $\mathcal{B}$ is a polynomial of degree $2\deg(V) - 3$. So, if V is a polynomial, then these considerations determine the algebraic curve in question, up to the choice of at most $3\deg(V) - 5$ undetermined coefficients.

Amazingly, the algebraic curve (42) is actually of much lower degree than might be expected. Indeed, some algebra shows that the curve has the simplified form

$$(45) \qquad w^3 - (V'(z))\, w^2 + \mathcal{C}_1(z)\, w - \mathcal{C}_0 = 0,$$

where $\mathcal{C}_1(z) = \mathrm{Poly}\left[V'(z)\left(f_1(z) + f_2(z)\right)\right] - a^2$ is a polynomial of degree $\deg(V) - 2$, and $\mathcal{C}_0$ is a polynomial of degree $\deg(V) - 1$. To see that this latter claim is true, note that one has the representation

$$(46) \qquad \mathcal{C}_0(z) = -a^2 V'(z) - \mathrm{Poly}\left[V'(z)\left(f_1^2 + 2a\left(f_2 - f_1\right) + f_2^2\right)\right] +$$

$$(47) \qquad + \mathrm{Poly}\left[\left(V'(z)\right)^2\left(f_1(z) + f_2(z)\right)\right] - V'(z)\mathrm{Poly}\left[V'(z)\left(f_1(z) + f_2(z)\right)\right],$$

which is easily seen to be a polynomial of degree $\deg(V) - 1$.

5. The existence problem.

We have shown that *if there exists* sets I_1 and I_2 and functions f_1 and f_2 satisfying the conditions of Lemma 3.1, then the functions f_1 and f_2 are related to the roots of a particular algebraic curve. On the other hand, for the purposes of asymptotic analysis of Riemann–Hilbert problems, what is really required is the *existence* itself: given external fields V_1 and V_2, one needs to know that functions f_1 and f_2, together with sets I_1 and I_2, exist so that the original Riemann–Hilbert problem 1.1 may be "regularized" into the form of Riemann–Hilbert problem 2.1, and subsequent asymptotic analysis may be carried out.

If one knew the algebraic curve explicitly, then by careful analysis one could prove that (i) sets I_1 and I_2 exist (they are the branch cuts of the algebraic functions determined by the algebraic equation), (ii) functions f_1 and f_2 exist (they are the actual algebraic functions) and (iii) all the required reality, and positivity/negativity conditions are satisfied (this is actually the most difficult part of the analysis).

Unfortunately, except in special Gaussian cases (see the next Section), the functions $\mathcal{C}_1$ and $\mathcal{C}_2$ are not completely determined by the considerations leading to, nor by the proof of, Theorem 3.2. For example, if V_1 and V_2 are polynomials, then one can prove (again, see the subsequent Section for concrete examples) that $\mathcal{C}_1$ and $\mathcal{C}_2$ are polynomials in z, whose leading coefficients are explicitly known. However, the explicit determination of the lower order terms does not follow from these considerations.

In other words, one has not shown existence of sets I_j and functions f_j unless one can prove that within the family of curves parametrized by the coefficients left undetermined by Theorem 3.2, there exists an algebraic curve which satisfies all the necessary conditions. We will formulate this as an existence problem.

Existence Problem 5.1. *Given polynomial (or more general) external fields V_1 and V_2, prove that there is an algebraic curve of the form (36) whose roots yield functions f_1 and f_2 as in (33)-(34) so that items A-D of the "ideal situation" described near the end of Section 2 (or their generalization to the case where the curves in question are not necessarily real) are satisfied.*

6. Existence theorems for quadratic and quartic external fields

The Gaussian case.

The Gaussian case corresponds to $V = x^2/2$. In this case, one may verify that

$$\mathcal{A}(z) = -1, \tag{48}$$
$$\mathcal{B}(z) = -z. \tag{49}$$

And so, we see that the desired algebraic equation is

$$w^3 - zw^2 - \left(a^2 - 1\right) w + a^2 z + a \left(2x_2 - 1\right) = 0. \tag{50}$$

We observe that in the Gaussian case, there are no undetermined coefficients at all, and, it is straightforward to verify (see, for example, [3], [1], for the case $x_1 = x_2 = 1/2$) that there is A so that for $a > A$, (50) possesses roots that yield functions g_j and sets I_j so that the conditions A-D hold true. (In the case $x_1 = x_2 = 1/2$, it is also known that for $a < A$, the equation (50) determines two piecewise analytic arcs as described in the third remark following the description of the conditions A-D in Section 3 [3, 1]).

The quartic case, with $x_1 = x_2 = 1/2$. The quartic case corresponds to $V(x) = x^4/4$. In this case, one may verify that

$$\mathcal{C}_1 = z^2 + c_1 z + c_0, \tag{51}$$
$$\mathcal{C}_0 = -a^2 z^3 + \hat{c}_2 z^2 + \hat{c}_1 z + \hat{c}_0. \tag{52}$$

If we consider the special symmetric case in which $x_1 = x_2 = 1/2$, then there is an additional symmetry satisfied by the functions f_1 and f_2, namely that

$$f_1(z) = -f_2(-z).$$

(This can be seen by considering $F_1(z) = -f_2(-z)$ and $F_2(z) = -f_1(-z)$, showing that F_1 and F_2 satisfy the same boundary value relations (20)-(21) as f_1 and f_2, and then using Theorem 3.1.)

This symmetry in turn implies that $\mathcal{C}_0$ is an odd polynomial, and $\mathcal{C}_1$ is an even polynomial, and hence the algebraic curve becomes

$$w^3 - z^3 w^2 + \left(z^2 + \alpha\right) w + a^2 z^3 + \beta z = 0, \tag{53}$$

and there are only two undetermined coefficients, α and β.

In order to arrive at a problem of the form of Riemann–Hilbert problem 2.1, the existence of sets I_1 and I_2, along with functions f_1 and f_2 must be established, as described in Section 3. As a demonstration that this is possible, we have the following result.

Theorem 6.1. *There is $A > 0$ so that for all $a > A$, there exists coefficients $\alpha(a)$ and $\beta(a)$ so that the following properties, pertaining to equation (53) with $\alpha = \alpha(a)$ and $\beta = \beta(a)$, are true.*

(1) *The three solutions $\{r_j(z)\}_{j=1}^3$ of the equation (53), determined uniquely as algebraic functions of z for $z \in \mathbb{C}$ that possess standard meromorphic Laurent expansions at ∞, satisfy the following*

9

asymptotic behavior for $z \to \infty$:

$$(54) \qquad r_3 = z^3 - \frac{1}{z} + \mathcal{O}\left(z^{-2}\right),$$

$$(55) \qquad r_2 = a + \frac{1}{2z} + \mathcal{O}\left(z^{-2}\right),$$

$$(56) \qquad r_1 = -a + \frac{1}{2z} + \mathcal{O}\left(z^{-2}\right).$$

(2) *The functions f_1 and f_2, defined by*

$$(57) \qquad f_1(z) = r_1(z) + a, \qquad f_2(z) = r_2(z) - a,$$

possess the following asymptotic expansions for $z \to \infty$:

$$(58) \qquad f_j = \frac{1}{2z} + \mathcal{O}\left(z^{-2}\right), \quad j = 1, 2.$$

(3) *There are two real positive real numbers $\gamma_1 = \gamma_1(a)$ and $\gamma_2 = \gamma_2(a)$ so that the function f_2 may be taken to be analytic in $\mathbb{C} \setminus I_2$ where $I_2 = [\gamma_1, \gamma_2]$.*

(4) *The function f_1 may be taken to be analytic in $\mathbb{C} \setminus I_1$ where $I_1 = [-\gamma_2, -\gamma_1]$.*

(5) *The following boundary relations hold true:*

$$(59) \qquad \left(f_1^- + f_1^+\right) + f_2 - V_1' = 0, \quad z \in I_1,$$

$$(60) \qquad \left(f_2^- + f_2^+\right) + f_1 - V_2' = 0, \quad z \in I_2.$$

(6) *For $z \in I_j$, we have*

$$(61) \qquad \frac{1}{i} \left(f_j^+(z) - f_j^-(z)\right) < 0.$$

(7) *For $z \in \mathbb{R} \setminus I_1$, we have*

$$(62) \qquad \int_{-\gamma_2}^{z} \mathrm{Re} \ \left(f_1^-(s) + f_1^+(s) + f_2^-(s) - V_1'(s)\right) \ ds < 0,$$

and for $z \in \mathbb{R} \setminus I_2$, we have the analogous inequality:

$$(63) \qquad \int_{\gamma_2}^{z} \mathrm{Re} \ \left(f_2^-(s) + f_2^+(s) + f_1^-(s) - V_2'(s)\right) \ ds < 0,$$

Remark: We note that for the quartic case considered here, Theorem 6.1 solves the Existence Problem as described in Section 3, for all $a > A$.

Proof of (1) and (2): Straightforward asymptotic calculations for $z \to \infty$ (which are left to the diligent reader) show that the three roots $\{r_j\}_{j=1}^3$ have Laurent expansions of the form (54)-(56), and it is immediately clear that (58) holds true.

Proof of (3) and (4): Whether one calculates the resultant of the polynomial

$$p_3(z, w) = w^3 - z^3 w^2 + \left(z^2 + \alpha\right) w + a^2 z^3 + \beta z$$

with its derivative $(\partial/\partial w)p_3(z, w)$ or just solves the equation (53) directly, one arrives at the following 12th degree polynomial in the variable z whose roots are potential branch points for the roots $\{r_j(z)\}_{j=1}^3$:

$$(64) \quad \tilde{q}(z; \alpha, \beta, a) = 36a^2 z^{12} + 9(4\beta + 1)z^{10} + 9\left(2\alpha - 18a^2\right)z^8 + 9\left(-27a^4 - 18\alpha a^2 + \alpha^2 - 18\beta - 4\right)z^6 +$$
$$+9\left(-54\beta a^2 - 12\alpha - 18\alpha\beta\right)z^4 + 9\left(-12\alpha^2 - 27\beta^2\right)z^2 - 36\alpha^3.$$

The polynomial $\tilde{q}$ is even and calculations are greatly simplified by considering

$$(65) \quad q(t) = \tilde{q}(\sqrt{t}) = 36a^2 t^6 + 9(4\beta + 1)t^5 + 9\left(2\alpha - 18a^2\right)t^4 + 9\left(-27a^4 - 18\alpha a^2 + \alpha^2 - 18\beta - 4\right)t^3 +$$
$$+9\left(-54\beta a^2 - 12\alpha - 18\alpha\beta\right)t^2 + 9\left(-12\alpha^2 - 27\beta^2\right)t - 36\alpha^3.$$

The goal is to find (α, β) so that this polynomial has exactly two positive simple roots and two double roots (that will turn out to be complex). We consider the resultant of $q(t)$ with its derivative $q'(t)$:

$$(66) \qquad \mathrm{Resultant}\left[q(t), q'(t)\right] = B_1(\alpha, \beta, a)B_2(\alpha, \beta, a),$$

where

(67) $\quad B_1(\alpha, \beta, a) = 729\alpha a^6 + 243\left(\alpha^2 - 3\beta\right)a^4 + 27\alpha\left(\alpha^2 - 15\beta - 1\right)a^2 + \alpha^4 + 27\beta(3\beta + 1)^2 - 36\alpha^2\beta,$

(68) $\quad B_2(\alpha, \beta, a) = 729\alpha^3 a^{10} + 729\left(3\alpha^4 - \beta\alpha^2 - \beta^3\right)a^8 + 27\left(81\alpha^5 + (8 - 171\beta)\alpha^3 - 72\beta^3\alpha\right)a^6 +$

$$+27\left(27\alpha^6 - 20(9\beta + 1)\alpha^4 - 2\beta\left(28\beta^2 - 47\beta + 4\right)\alpha^2 + 8\beta^3(12\beta - 1)\right)a^4 +$$

$$+\left(-27(36\beta + 1)\alpha^5 + 4\left(216\beta^3 + 810\beta^2 + 189\beta + 4\right)\alpha^3 + 576\beta^3(4\beta + 1)\alpha\right)a^2 +$$

$$-\beta(4\beta + 1)^2\left(-27\alpha^4 + 8(9\beta + 2)\alpha^2 + 16\beta^2(4\beta + 1)\right).$$

Lemma 6.1. *There is A so that for all $a > A$, the following statement is true: For real α and β, there is an isolated zero of $B_2(\alpha, \beta)$, which is a local maximum, occurring at $(\alpha^*(a), \beta^*(a))$. The quantities $\alpha^*(a)$ and $\beta^*(a)$ possess the following expansions:*

(69)
$$\alpha^*(a) = a^2\left(-1 + a^{-4/3} + \left(\frac{1}{27}\right)a^{-12/3} + a^{-16/3}\sum_{j=0}^{\infty}\alpha_j\left(a^{-4/3}\right)^j\right)$$

(70)
$$\beta^*(a) = a^{4/3}\left(1 - \left(\frac{1}{3}\right)a^{-4/3} + a^{-12/3}\sum_{j=0}^{\infty}\beta_j\left(a^{-4/3}\right)^j\right)$$

Proof

For the polynomial $B_2(\alpha, \beta)$ to have a local maximum with value 0, it must vanish together with its first partial derivatives. To prove this, we first observe that

i the resultant of B_2 with $\partial_\alpha B_2$ (eliminating β) coincides, up to a constant factor, with the resultant of B_2 with $\partial_\beta B_2$:

(71)
$$\text{Resultant}\left[B_2, \partial_\alpha B_2\right](\alpha) = \frac{a^4}{2048}\text{Resultant}\left[B_2, \partial_\beta B_2\right](\alpha).$$

ii Moreover, there is a unique value of α (called α_*) with the asymptotic expansion (69) such that the above resultants vanish. There are twelve other roots (of varying multiplicities), which remain bounded away from this root by a distance of $\mathcal{O}\left(a^{2/3}\right)$.

iii In similar fashion, the resultants of B_2 with its first partials, eliminating α instead of β, are also equivalent modulo a constant factor:

(72)
$$\text{Resultant}\left[B_2, \partial_\alpha B_2\right](\beta) = -a^4\text{Resultant}\left[B_2, \partial_\beta B_2\right](\beta). \quad).$$

iv Furthermore, there is a unique value of β (called β_*) with the asymptotic expansion (70) so that *these* two resultants vanish. There are 12 other roots (of varying multiplicities), which remain bounded away from this root by a distance of $\mathcal{O}\left(a^{4/3}\right)$ (independent of a).

The assertions ii and iv above may be verified by computing the large a behavior of these resultants.

Next one considers the Sylvester matrix associated to $\text{Resultant}\left[B_2, \partial_\beta B_2\right](\alpha)$ (i.e. the matrix whose determinant is this resultant), and shows that for $\alpha = \alpha_*$, there is an eigenvector of the form

(73)
$$\left(\beta(\alpha_*)^{10}, \beta(\alpha_*)^9, \ldots, \beta(\alpha_*), 1\right)^t$$

with $\beta(\alpha_*)$ possessing the first few terms of the asymptotic expansion (70). Since $\beta(\alpha_*)$ must be a root of the polynomial $\text{Resultant}\left[B_2, \partial_\beta B_2\right](\beta)$, it follows that $\beta(\alpha_*) = \beta_*$.

It now follows that B_2, along with its first partial derivatives, vanishes for $(\alpha, \beta) = (\alpha_*, \beta_*)$.

Again using the fact that a may be taken large, one may verify that for (α, β) near (α_*, β_*), the surface defined by $B_2(\alpha, \beta)$ satisfies

(74)
$$\text{Hess}\left(B_2\right)(\alpha_*, \beta_*) = -4782969\, a^{80/3}\left(1 + \mathcal{O}\left(a^{-4/3}\right)\right).$$

That an A exists so that for all $a > A$ the above holds true is a straightforward exercise in perturbation theory. $\square$

Remark: From this point on, we shall take $a > A$, and $\alpha = \alpha_*$, $\beta = \beta_*$. However, to avoid cumbersome notation, we will usually not emphasize this. For example, $q(t)$ will now refer to $q(t; \alpha = \alpha_*, \beta = \beta_*)$, etc.

A consequence of Lemma 6.1 is that the polynomial $q(t)$ defined in (65) possesses two complex double roots. One sees this by first carrying out an asymptotic expansion of $q(t)$ for $a \to \infty$, using the asymptotic expansions (69) and (70), to see that $q(t)$ possess the following asymptotic expansion:

$$(75) \qquad a^{-6}\, q(a^{2/3}\zeta) = 36(\zeta - 1)^2(\zeta^2 + \zeta + 1)^2 + 36\, a^{-4/3}\left(1 + \zeta + \zeta^2\right)\left(-3 - 6\,\zeta^2 + \zeta^3\right) +$$
$$-3\, a^{-8/3}\left(-36 + 9\,\zeta - 18\,\zeta^2 + 51\,\zeta^3 - 6\,\zeta^4 + \zeta^5\right) +$$
$$+18\, a^{-12/3}\left(-2 + \zeta\right)\left(1 - \zeta + \zeta^2\right) - 27a^{-16/3}\zeta.$$

Perturbative calculations then show that for a sufficiently large, $q(a^{2/3}\zeta)$ possesses exactly two real roots. These are simple and positive (near $\zeta = 1$). In fact, the reader may verify that the two roots, which we define to be γ_1^2 and γ_2^2 (where γ_1 and γ_2 are two positive numbers) satisfy:

$$(76) \qquad \gamma_2^2 = a^{2/3}\left[1 + 2\sqrt{\frac{2}{3}}a^{-2/3} + \cdots\right]$$

$$(77) \qquad \gamma_1^2 = a^{2/3}\left[1 - 2\sqrt{\frac{2}{3}}a^{-2/3} + \cdots\right]$$

Therefore $q(t)$ has four complex roots, appearing in complex conjugate pairs. Since $q(t)$ must have at least one double root, there must be two double roots. Therefore $q(t)$ factors as follows:

$$(78) \qquad q(t) = 36a^2(t - \gamma_1(a)^2)(t - \gamma_2(a)^2)\left[(t - \lambda_*)(t - \overline{\lambda_*})\right]^2,$$

with $\mathrm{Imag}\,(\lambda_*) > 0$.

Now if one solves the cubic (53), one finds

$$(79) \qquad w = \frac{1}{3}z^3 + \left(\frac{-R(z) + 3^{-5/2}\sqrt{-q(z^2)}}{2}\right)^{1/3} + \frac{H(z)}{\left(\frac{-R(z)+3^{-5/2}\sqrt{-q(z^2)}}{2}\right)^{1/3}},$$

where

$$(80) \qquad R(z) = -\frac{2}{27}z^9 + \frac{1}{3}z^5 + \frac{\alpha}{3}z^3 + a^2 z^3 + \beta z,$$

$$(81) \qquad H(z) = \frac{1}{9}z^6 - \frac{1}{3}z^2 - \frac{1}{3}\alpha,$$

and the cube roots appearing in (79) are all the same choice of branch, but not necessarily the principal branch (the three different choices yield different solutions to the cubic (53)). Similarly, the square roots appearing in (79) are all the same choice of branch, not necessarily the principal branch. (However, different choices of the branch of the square root do not yield different solutions.)

Remark: The representation (79), although concise, is at present only useful for local analyses since the choice of branches, usually a delicate matter, has been avoided. We will use this representation together with local analysis near zeros of $q(z^2)$ to describe the possible branch points of the solutions to (53), and later we will make a particular choice of branches to carry out a global analysis.

We already know that the only potential branching points where a solution to (53) could fail to be analytic are the simple roots of $q(z^2)$. Yet another asymptotic analysis for large a shows that $R(z)$ does not vanish at any of these roots, and so we conclude that there are only four branch points, namely $\pm\gamma_1(a)$ and $\pm\gamma_2(a)$, and at each of these points, the solutions to (53) exhibit square-root singularities. (So, for example, in a vicinity of γ_1, the solutions are analytic functions of $\sqrt{z - \gamma_1}$.)

At each of the four branch points, the square root appearing in (79) vanishes, while $R(z)$ does not, and so a local analysis in a vicinity of each branch point can be obtained by choosing a particular branch for the square root (as will be done below), and then, for each distinct choice of branch of the cube-root, use the

12

Taylor expansions

$$(82) \left(-R + 3^{-5/2}\sqrt{-q(z^2)}\right)^{1/3} = \sqrt[3]{\left|\frac{-R}{2}\right|}e^{i\phi/3}\left\{1 - \frac{1}{3}\left(\frac{3^{-5/2}\sqrt{-q(z^2)}}{R(z)}\right) - \frac{1}{9}\left(\frac{3^{-5/2}\sqrt{-q(z^2)}}{R(z)}\right)^2 + \right.$$

$$\left. - \frac{5}{81}\left(\frac{3^{-5/2}\sqrt{-q(z^2)}}{R(z)}\right)^3 - \frac{10}{243}\left(\frac{3^{-5/2}\sqrt{-q(z^2)}}{R(z)}\right)^4 + \cdots\right\},$$

$$(83) \left(-R - 3^{-5/2}\sqrt{-q(z^2)}\right)^{1/3} = \sqrt[3]{\left|\frac{-R}{2}\right|}e^{i\phi/3}\left\{1 + \frac{1}{3}\left(\frac{3^{-5/2}\sqrt{-q(z^2)}}{R(z)}\right) - \frac{1}{9}\left(\frac{3^{-5/2}\sqrt{-q(z^2)}}{R(z)}\right)^2 + \right.$$

$$\left. + \frac{5}{81}\left(\frac{3^{-5/2}\sqrt{-q(z^2)}}{R(z)}\right)^3 - \frac{10}{243}\left(\frac{3^{-5/2}\sqrt{-q(z^2)}}{R(z)}\right)^4 + \cdots\right\},$$

where the phase $\phi = \phi(z)$ is the complex argument of $-R(z)$. Different choices of this phase yield the three different solutions to (53).

However, before doing this, it is advantageous to eliminate the radical from the denominator of (79):

$$(84) \qquad \frac{H(z)}{\left(\frac{-R(z)+3^{-5/2}\sqrt{-q(z^2)}}{2}\right)^{1/3}} = e^{-2i\phi/3}\left(\frac{-R(z) - 3^{-5/2}\sqrt{-q(z^2)}}{2}\right)^{1/3}.$$

Now using the convergent Taylor expansions (82) and (83) in (79), the reader may verify (with some sweat) that exactly one of the three possible branches of the cube-root is analytic in a neighborhood of each branch point. Indeed for that one choice of cube-root, all odd powers of the square-root happen to vanish! For example, for z real and near γ_2, it turns out that $R(z) > 0$ (see Lemma 6.2 below) and hence one must choose ϕ to be an odd multiple of π. The three choices $-\pi, \pi, 3\pi$ will yield all possible solutions to the cubic equation (53). Now using the Taylor expansions (82) and (83), one finds

$$(85) \;\; w(z) - \frac{1}{3}z^3 = \left(\frac{-R(z)+3^{-5/2}\sqrt{-q(z^2)}}{2}\right)^{1/3} + \frac{H(z)}{\left(\frac{-R(z)+3^{-5/2}\sqrt{-q(z^2)}}{2}\right)^{1/3}} =$$

$$\sqrt[3]{\left|\frac{-R}{2}\right|}\left(\sum_{j=0,\; j\text{ even}}^{\infty} 2c_j\left(\frac{3^{-5/2}\sqrt{-q(z^2)}}{R(z)}\right)^j \cos\left(\phi/3\right) \;-\; \sum_{j=1,\; j\text{ odd}}^{\infty} 2i\tilde{c}_j\left(\frac{3^{-5/2}\sqrt{-q(z^2)}}{R(z)}\right)^j \sin\left(\phi/3\right)\right),$$

with $c_0 = 1$, $\tilde{c}_1 = 1/3$, and clearly the choice $\phi = 3\pi$ is the only one for which all odd powers of the square root vanish, yielding a solution (79) to (53) which is analytic in a neighborhood of γ_2, while the remaining two are not.

Remark: Observe that for the other two choices of the angle ϕ, namely $-\pi$ and π, the two corresponding roots coincide at $z = \gamma_2$. These two roots are discontinuous across $I_2 = (\gamma_1, \gamma_2)$. It follows that the boundary value on one side of I_2 of one of these two roots must coincide with the opposing boundary value of the *other* root. Similarly, for the two roots that are not analytic across I_1, their boundary values from opposing sides coincide, and at the endpoints of the interval I_1, those two roots are equal.

Global analyticity properties

In order to describe the global analyticity properties of the solutions to (53), we will make the following choice for the branch of the square root appearing in (79):

$$(86) \qquad \sqrt{-q(z^2)} := 6ia\left[(z^2 - \lambda_*)(z^2 - \overline{\lambda_*})\right]\sqrt{z - \gamma_1(a)}\sqrt{z - \gamma_2(a)}\sqrt{z + \gamma_1(a)}\sqrt{z + \gamma_2(a)}$$

where now $\sqrt{\cdot}$ is the principal branch. The function appearing in (86) is analytic in $\mathbb{C}\setminus\{[-\gamma_2, -\gamma_1] \cup [\gamma_1, \gamma_2]\}$.

Using this choice of branch for $\sqrt{-q(z^2)}$ in (79), along with the fact that $R(z)$ is strictly positive on I_2, it follows that the one solution which is analytic in a neighborhood of γ_2 can be continued to a fixed

neighborhood of the interval $[\gamma_1, \gamma_2]$. This solution is then analytic for $z \in \mathbb{C} \backslash [-\gamma_2, -\gamma_1]$ (of course, it cannot be entire, and so it must have a branch cut, which can only be across $(-\gamma_2, -\gamma_1)$). This solution shall be referred to as $r_1(z; a)$.

A similar argument shows that there is another solution, which will be called $r_2(z; a)$, that is analytic in $\mathbb{C} \backslash [\gamma_1, \gamma_2]$.

Remark: In order to complete the proof of (3) and (4) of Theorem 6.1, we will show that the solutions $r_1(z; a)$ and $r_2(z; a)$ possess the $z \to \infty$ asymptotic expansions (56) and (55), respectively. We will only consider the case of $r_1(z; a)$, as the proof that the solution $r_2(z; a)$ possesses the expansion (55) is entirely similar.

The starting point for this global analysis is the signature of the functions $R(z)$, $H(z)$, and $q(z^2)$ in a vicinity of $\gamma_2(a)$, and we are once again aided by the fact that a may be taken large. Indeed, we have the following lemma.

Lemma 6.2. *There is $A > 0$ so that for all $a > A$, the following statements are true.*
- *There is $\eta(a) > \gamma_2(a)$ so that the function $R(z)$ is positive on the interval $(0, \eta)$, and is negative on $(\eta(a), \infty)$.*
- *The function $H(z)$ is strictly positive for all $z > 0$*

Proof

Using the rescaling $z \mapsto a^{1/3}\zeta$, we find

$$(87) \qquad a^{-3}R(a^{1/3}\zeta) = p(\zeta) + a^{-4/3}\hat{p}(\zeta),$$

where

$$(88) \qquad p(\zeta) = \frac{1}{27}\zeta^3 \left(27 + 9 \left(\frac{\alpha_*}{a^2} \right) - 2\,\zeta^6 \right)$$

$$(89) \qquad \hat{p}(\zeta) = \left(\left(\frac{\beta_*}{a^{4/3}} \right) \zeta + \frac{\zeta^5}{3} \right).$$

Now α_* and β_* possess asymptotic expansions (cf. (69),(70)) so that the coefficients $\frac{\alpha_*}{a^2}$ and $\frac{\beta_*}{a^{4/3}}$ in (88) and (89) are uniformly bounded, and

$$(90) \qquad \frac{\alpha_*}{a^2} \to -1, \qquad \frac{\beta_*}{a^{4/3}} \to 1.$$

The reader may verify that the leading order term, $p(\zeta)$, possesses one positive real root which approaches $3^{1/3}$. Moreover, $p(\zeta)$ is positive to the left of this root, and negative to the right.

Next we consider ζ in a fixed interval of the form $[\frac{\gamma_2}{a^{1/3}}, X]$. The function $\hat{p}(\zeta)$ is uniformly bounded on this interval, and so the function $a^{-3}R(a^{1/3}\zeta)$ can only possess roots in a vicinity of the root of $p(\zeta)$. However, this is a simple root of $p(\zeta)$, and the derivative $(d/d\zeta)a^{-3}R(a^{1/3}\zeta)$ is strictly negative in a vicinity of this root, guaranteeing that $a^{-3}R(a^{1/3}\zeta)$ possesses exactly one simple root in the interval $[\frac{\gamma_2}{a^{1/3}}, X]$.

Lastly one considers the behavior for ζ large, where the function $p(\zeta)$ dominates the function $\hat{p}(\zeta)$, and one may then choose X large enough (but fixed) so that $R(a^{1/3}\zeta)$ possesses no roots beyond X, and this finishes the proof of the first claim of the Lemma.

The second claim can be established by similar reasoning, using the asymptotic representation

$$a^{-2}H(a^{1/3}\zeta) = \frac{1}{9}\left(\zeta^6 - 3 \left(\frac{\alpha_*}{a^2} \right) \right) - a^{-4/3}\left(\frac{1}{3}\,\zeta^2 \right)$$

It is straightforward to verify that the leading order term on the right hand side is strictly positive for $\zeta > \gamma_2/a^{1/3}$, and the remaining details of the proof of the second claim are left to the reader. $\qquad \square$.

Remark: For the sake of completeness, we note that the function $\sqrt{-q(z^2)}$ is real and negative for $z \in (\gamma_1(a), \gamma_2(a))$, and positive imaginary for $z > \gamma_2$.

Using Lemma 6.2, it follows that as z traverses the real axis, from γ_2 to $+\infty$, the quantity $-R(z) + 3^{-5/2}\sqrt{-q(z^2)}$ starts out on the negative real axis, enters the second quadrant, exiting the second quadrant by crossing the positive imaginary axis, and then diverges to ∞ in the first quadrant.

14

In the text following (85), we have shown that for $z \in \mathbb{R}$ and near γ_2, the appropriate choice of cube root yielding a function that is analytic is determined by $\phi = 3\pi$. As z traverses the real axis from γ_2 to $+\infty$, this angle must converge to 2π, and hence we have determined the behavior of the root $r_1(z)$ to be as follows:

$$
\text{(91)} \qquad r_1(z) = \frac{1}{3}z^3 + \frac{1}{3}z^3 e^{2i\pi/3}\left(1 + \frac{ia\sqrt{3}}{z^3} + -\frac{3}{2z^4} + \mathcal{O}\left(\frac{1}{z^5}\right)\right) +
$$

$$
+\frac{1}{3}z^3 e^{-2i\pi/3}\left(1 - \frac{ia\sqrt{3}}{z^3} - \frac{3}{2z^4} + \mathcal{O}\left(\frac{1}{z^5}\right)\right)
$$

$$
= -a + \frac{1/2}{z} + \mathcal{O}\left(\frac{1}{z^2}\right),
$$

which establishes (56).

We will leave to the motivated reader the proof that the root r_2, which is analytic in $\mathbb{C}\setminus[\gamma_1,\gamma_2]$, possesses the asymptotic expansion (55) as $z \to \infty$.

This completes the proof of (3) and (4) of Theorem 6.1.

Proof of (5) of Theorem 6.1: To see that the boundary relation (59) holds true, one first observes that for $z \in I_1$, both boundary values $r_1^{\pm}$ are roots of the cubic (53). Since they are not both the same root, and since r_2 is analytic in a neighborhood of I_1, it follows that there must be a relation between $r_1^{\pm}$ and $r_3^{\pm}$. The reader may verify that this monodromy relation is

$$
r_1^{\pm} = r_3^{\mp} \quad \text{for } z \in I_1.
$$

It is then straightforward to verify that

$$
\left(f_1^- + f_1^+\right) + f_2 = r_1^+ + r_1^- + r_2 + a = r_1^+ + r_2 + r_3^+ + a = z^3 + a,
$$

since the sum of the roots of the cubic (53) must be z^3. The boundary relation (60) follows by very similar reasoning, the difference being that on I_2, the boundary values $r_2^{\pm}$ are related to the boundary values of r_3 as follows: $r_2^{\pm} = r_3^{\mp}$.

Proof of (6): The inequalities (61) may be seen to be true by one of the following two approaches. We will consider $j = 1$, as the considerations for $j = 2$ are entirely similar.

(1) Using the expansions (82) and (83) (with suitable choice of the angle ϕ), valid over the entire interval I_1, one finds that the difference between the boundary values from the upper and lower half planes is positive imaginary.

(2) One shows that the two roots r_1^+ and r_1^- are complex conjugates of each other, and that r_1^+ has positive imaginary part.

Proof of (7): We now prove that the inequality (62) is true. First, recall that on I_1, (59) holds true. Next observe that for $s \in \mathbb{R}\setminus I_1$,

$$
\text{(92)} \qquad f_1^-(s) + f_1^+(s) + f_2^-(s) - V_1'(s) = 2f_1(s) + f_2^-(s) - s^3 - a = r_1(s) - r_3^-(s).
$$

It turns out that one must verify the inequality (62) separately on each of the intervals $(-\infty, -\gamma_2)$, $(-\gamma_1, \gamma_1)$, I_2, and (γ_2, ∞).

Local analysis near the branch points $-\gamma_2$, $-\gamma_1$, and γ_2 show that on each of the intervals $(-\infty, -\gamma_2)$, $(-\gamma_1, \gamma_1)$ and (γ_2, ∞) there are three real roots. Since the only places where any two roots can coincide are the branch points $\pm\gamma_1$ and $\pm\gamma_2$, the quantity (92) is of one sign for s in each of $(-\infty, -\gamma_2)$, $(-\gamma_1, \gamma_1)$ and (γ_2, ∞). The signature in each of these intervals can be determined by a (by now) straightforward local analysis in a vicinity of each branch point.

For example, for $s = -\gamma_1$, the two roots r_1 and r_3 coincide, and the following local expansions hold true:

$$(93) \qquad r_1(s) - \frac{1}{3}s^3 =$$

$$\sqrt[3]{\left|\frac{-R}{2}\right|}\left[\cos\left(2\pi/3\right)\left(1 + \mathcal{O}\left(\left|q(s^2)\right|\right)\right) - \frac{2i}{3}\left(\frac{3^{-5/2}\sqrt{-q(s^2)}}{R(s)}\right)\sin\left(2\pi/3\right)\left(1 + \mathcal{O}\left(\left|q(s^2)\right|\right)\right)\right],$$

$$(94) \qquad r_3(s) - \frac{1}{3}s^3 =$$

$$\sqrt[3]{\left|\frac{-R}{2}\right|}\left[\cos\left(4\pi/3\right)\left(1 + \mathcal{O}\left(\left|q(s^2)\right|\right)\right) - \frac{2i}{3}\left(\frac{3^{-5/2}\sqrt{-q(s^2)}}{R(s)}\right)\sin\left(4\pi/3\right)\left(1 + \mathcal{O}\left(\left|q(s^2)\right|\right)\right)\right].$$

Now for s real, $s > -\gamma_1$, and s near $-\gamma_1$, $\sqrt{-q(s^2)}$ is a number on the negative imaginary axis. Therefore

$$(95) \qquad r_1(s) - r_3(s) = -\sqrt[3]{\left|\frac{-R}{2}\right|}\frac{4i}{3}\left(\frac{3^{-5/2}\sqrt{-q(s^2)}}{R(s)}\right)\sin\left(2\pi/3\right) + \mathcal{O}\left(\left|q(s^2)\right|\right) < 0,$$

and then it follows that the inequality (62) is true for $s \in (-\gamma_1, \gamma_1)$. Similar analysis shows that the same inequality is true on $(-\infty, -\gamma_2)$, and we will leave the details of those calculations to the reader.

For z in the interval I_2, the root r_1 remains real, but the root r_3 (either the "+" or the "−" boundary value) is not real, and the quantity $\mathrm{Re}\left(r_1(z) - r_3^{(-)}(z)\right)$ may change signs without two roots coinciding.

However, recalling the form (79) of the roots of the cubic (53), along with the useful relationship (84), we may represent the roots r_j as follows:

$$(96) \qquad r_j = \frac{1}{3}z^3 + \left(\frac{-R(z) + 3^{-5/2}\sqrt{-q(z^2)}}{2}\right)^{1/3} + e^{-2i\phi_j/3}\left(\frac{-R(z) - 3^{-5/2}\sqrt{-q(z^2)}}{2}\right)^{1/3},$$

where the choice of cube-root, and the phase ϕ_j, depends upon which root is being considered.

We will need the following properties:

- The function $\sqrt{-q(z^2)}$ is real and negative for $z \in (\gamma_1(a), \gamma_2(a))$.
- The function $R(z)$ is strictly positive for $z \in (\gamma_1(a), \gamma_2(a))$.
- The quantities $-R(z) + 3^{-5/2}\sqrt{-q(z^2)}$ and $-R(z) - 3^{-5/2}\sqrt{-q(z^2)}$ are nonzero for $z \in [\gamma_1(a), \gamma_2(a)]$, and consequently they are both strictly negative for all $z \in [\gamma_1(a), \gamma_2(a)]$.

With the above properties, and the choice $\phi_1 = 3\pi$ (see the text following (85)), we have

$$(97) \qquad \left(r_1 - r_3^-\right) = e^{i\pi}\left|\frac{-R(z) + 3^{-5/2}\sqrt{-q(z^2)}}{2}\right|^{1/3} + e^{i\pi}\left|\frac{-R(z) - 3^{-5/2}\sqrt{-q(z^2)}}{2}\right|^{1/3} +$$

$$- \left(e^{i\phi_3/3}\left|\frac{-R(z) + 3^{-5/2}\sqrt{-q(z^2)}}{2}\right|^{1/3} + e^{-2i\phi_j/3}e^{i\phi_j/3}\left|\frac{-R(z) - 3^{-5/2}\sqrt{-q(z^2)}}{2}\right|^{1/3}\right)$$

$$= \left(e^{i\pi} - e^{i\phi_3/3}\right)\left|\frac{-R(z) + 3^{-5/2}\sqrt{-q(z^2)}}{2}\right|^{1/3} + \left(e^{i\pi} - e^{-i\phi_3/3}\right)\left|\frac{-R(z) - 3^{-5/2}\sqrt{-q(z^2)}}{2}\right|^{1/3},$$

and therefore we have shown that $\mathrm{Re}\left(r_1 - r_3^-\right) < 0$ for all $z \in [\gamma_1, \gamma_2]$, which completes the proof that (62) is true for $z \in \gamma_1, \gamma_2)$.

Lastly, since $\mathrm{Re}\left(r_1 - r_3^-\right) < 0$ for $z \in [\gamma_1, \gamma_2]$, and since all three roots are distinct for all $z \in (\gamma_2, \infty)$, it follows that this inequality remains true in (γ_2, ∞) as well, and hence we have completed the proof that (62) holds true for all $z \in \mathbb{R} \setminus I_1$.

The proof that (63) holds true follows by very similar arguments, and we will omit these details here.

Having completed the proof of (7) of Theorem 6.1, we have completed the proof of the entire Theorem.

$\square$

Remark: We note that at each of the four endpoints, the roots r_1 and r_2 possess square root singularities. This implies that in addition to verifying all of the equalities and inequalities of the ideal situation A-D, conditions E1 and E2 also hold true.

7. Complete analysis of the Riemann–Hilbert problem 1.1 in the "one-cut" case

The starting point of this section is the Riemann–Hilbert problem (12). In this section we will show that *if* the ideal situation described in section 3 is achieved (including the "Airy conditions" E1 and E2), and if each of the sets $I_j, j = 1, 2$ is a single interval, then one may obtain a complete asymptotic description of the solution of the Riemann–Hilbert problem (12). Bleher and Kuijlaars [3] have carried out a Riemann–Hilbert analysis for the Gaussian case, and their work may be applied directly to the present more general situation. Therefore, we will only present a summary of the relevant transformations. Of course, it should be possible to generalize the work of this section to the case of arbitrary (finite) numbers of subintervals comprising the sets I_1 and I_2, but this is putting the cart before the horse as there is at present no existence theorem in the multi-cut case. In what follows we will let the endpoints of the intervals be defined as follows:

$$(98) \qquad I_1 = (\gamma_1, \delta_1), \qquad I_2 = (\gamma_2, \delta_2),$$

and we will use the following particular choices of antiderivatives of f_1 and f_2 to define g_1 and g_2:

$$(99) \qquad g_1 = \frac{n}{n_1} \int_{\delta_1}^z f_1(s)ds - c_1 \ ,$$

$$(100) \qquad g_2 = \frac{n}{n_2} \int_{\delta_2}^z f_2(s)ds - c_2 \ ,$$

where in (99) the contour is chosen in the cut plane $\mathbb{C} \setminus (-\infty, \delta_1)$, and in (100) the contour is taken in $\mathbb{C} \setminus (-\infty, \delta_2)$, and the constants c_1 and c_2 are chosen so that (9) holds true, namely

$$(101) \qquad c_1 = \frac{1}{2} \log \delta_1 \ + \int_{\delta_1}^{+\infty} \left(f_1(s) - \frac{1}{2s} \right) ds,$$

$$(102) \qquad c_2 = \frac{1}{2} \log \delta_2 \ + \int_{\delta_2}^{+\infty} \left(f_2(s) + \frac{1}{2s} \right) ds.$$

(Of course, then g_0 is determined via (10).)

Remark: In what follows, we will assume a certain type of *genericity* in the following sense. We are considering a sequence of triples, (n_1, n_2, n) so that $n = n_1 + n_2$ and $n_j/n \to x_j$. The implicit characterization of the functions g_1 and g_2 in terms of the equation $E(w, z) = 0$ with $E(w, z)$ defined in (36) depends on x_1 and x_2, and we will assume that the "ideal situation" holds true not just for x_1 and x_2, but for all values of (x'_1, x'_2) near (x_1, x_2).

The starting point is to realize that, because all of the conditions of the "ideal situation" are satisfied, the jump matrix appearing in (12), namely

$$(103) \qquad V_B = \begin{pmatrix} e^{-n(g_0^+ - g_0^-)} & e^{ng_0^- + n_1 g_1^+ - nV_1 - n_1\ell_1} & e^{ng_0^- + n_2 g_2^+ - nV_2 - n_2\ell_2} \\ 0 & e^{n_1(g_1^+ - g_1^-)} & 0 \\ 0 & 0 & e^{n_2(g_2^+ - g_2^-)} \end{pmatrix},$$

takes on one of four forms, as follows.

$$(104) \qquad V_B(x) = \begin{pmatrix} 1 & e^{-nP_1(x)} & e^{-nP_2(x)} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{for } x \in (-\infty, \gamma_1) \cup (\delta_1, \gamma_2) \cup (\delta_2, \infty),$$

$$(105) \qquad V_B(x) = \begin{pmatrix} e^{-in_2\theta_2(x)} & e^{-nP_1(x)} & 1 \\ 0 & 1 & 0 \\ 0 & 0 & e^{in_2\theta_2(x)} \end{pmatrix}, \quad \text{for } x \in I_2$$

$$(106) \qquad V_B(x) = \begin{pmatrix} e^{-in_1\theta_1(x)} & 1 & e^{-nP_2(x)} \\ 0 & e^{in_1\theta_1(x)} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{for } x \in I_1,$$

where

- $P_1 = x_1(g_1^+ + g_1^-) + x_2 g_2 - V_1 - x_1 \ell_1$ is strictly negative on $(-\infty, \gamma_1) \cup (\delta_1, \infty)$,
- $P_2 = x_2(g_2^+ + g_2^-) + x_1 g_1 - V_2 - x_2 \ell_2$ is strictly negative on $(-\infty, \gamma_2) \cup (\delta_2, \infty)$,
- $\theta_2(x) = -i(g_2^+(x) - g_2^-(x))$ is strictly positive and strictly decreasing on I_2, and possesses an analytic continuation (also called θ_2) off the interval I_2,
- $\theta_1(x) = -i(g_1^+(x) - g_1^-(x))$ is strictly positive and strictly decreasing on I_1, and possesses an analytic continuation (also called θ_1) off the interval I_1,

7.1. **Second transformation: Opening lenses.** Figure 1 below shows a decomposition of the plane into five regions, formed by defining four lens shaped regions surrounding the intervals I_1 and I_2. We will define the matrix D in each of these regions, as shown in Figure 1. (In the region exterior to the four lens shaped regions, $D = B$.) The matrix D is clearly piecewise analytic. Because it is related directly to the matrix B, it also solves a Riemann–Hilbert problem. However, we will not write this problem down, as it is somewhat auxiliary to our final goal. This definition is quite similar to the definitions (5.2), (5.4) and (5.6) of [3].
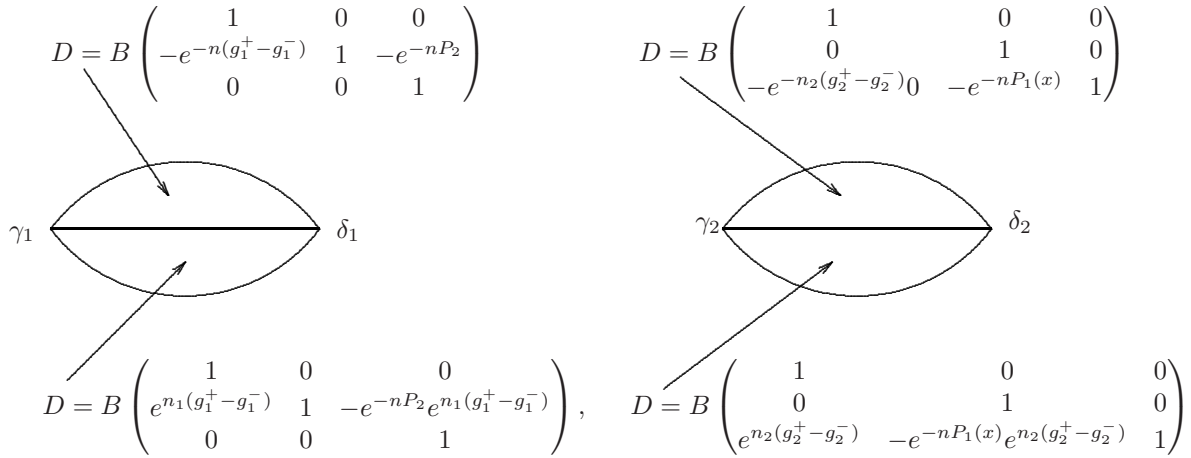
$$D = B \begin{pmatrix} 1 & 0 & 0 \\ -e^{-n(g_1^+ - g_1^-)} & 1 & -e^{-nP_2} \\ 0 & 0 & 1 \end{pmatrix} \qquad D = B \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -e^{-n_2(g_2^+ - g_2^-)} 0 & -e^{-nP_1(x)} & 1 \end{pmatrix}$$

$$\gamma_1 \qquad \delta_1 \qquad\qquad \gamma_2 \qquad \delta_2$$

$$D = B \begin{pmatrix} 1 & 0 & 0 \\ e^{n_1(g_1^+ - g_1^-)} & 1 & -e^{-nP_2}e^{n_1(g_1^+ - g_1^-)} \\ 0 & 0 & 1 \end{pmatrix}, \qquad D = B \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ e^{n_2(g_2^+ - g_2^-)} & -e^{-nP_1(x)}e^{n_2(g_2^+ - g_2^-)} & 1 \end{pmatrix}$$

Figure 1. The decomposition of the plane into 5 regions, the interior of the 4 lens shaped regions surrounding the intervals (γ_1, δ_1) and (γ_2, δ_2), and the one exterior region. The matrix D is defined as shown in each of the 4 bounded regions, and $D = B$ for z exterior to all four lens shaped regions.

Remark: In order to define the matrix D, one must extend, for $j = 1, 2$, the quantities $g_j^+ - g_j^-$ and P_j, for $j = 1, 2$ off the interval I_j. The justification for this uses the connection to the algebraic curve. For example,

$$(107) \qquad g_2^+(z) - g_2^-(z) = \frac{1}{x_2} \int_{\delta_2}^z f_2^+ - f_2^- \, dx = \int_{\delta_2}^z f_2^+ - f_0^+ \, dx = \int_{\delta_2}^z f_0^- - f_2^- \, dx,$$

which clearly demonstrates the local analyticity of $g_2^+ - g_2^-$. The other quantities are seen to be analytic by similar calculations.

The next step is to define a piecewise analytic matrix valued function $D_{approx}(z)$ which, it will be shown, is a globally uniform approximation to the matrix valued function D. We will require the following auxiliary functions. Set

$$(108) \qquad P_{\gamma_1} = r_1(\gamma_1), \qquad P_{\delta_1} = r_1(\delta_1),$$

$$(109) \qquad P_{\gamma_2} = r_2(\gamma_2), \qquad P_{\delta_2} = r_2(\delta_2).$$

$$(110)$$

Let Γ_1^+ denote the image of the $+$-side of the interval I_1 under the transformation $\xi(z) = r_0(z)$ (recall that $r_0(z)$ has branch cuts along I_1 and I_2), and let Γ_2 denote the image of the $+$-side of the interval I_2 under the same transformation $\xi(z) = r_0(z)$. The reader may verify that for each $j = 1, 2$, Γ_j^+ is an arc in the upper half plane connecting P_{γ_j} and P_{δ_j}.

18

Now we define three functions, $M_j(z)$ $(j = 1, 2, 3)$ as follows (cf. (6.9) of [3]).

$$(111) \qquad M_1(\xi) = \frac{\xi^2 - a^2}{\sqrt{(\xi - P_{\gamma_1})(\xi - P_{\delta_1})(\xi - P_{\gamma_2})(\xi - P_{\delta_2})}},$$

$$(112) \qquad M_2(\xi) = c_2 \frac{\xi + a}{\sqrt{(\xi - P_{\gamma_1})(\xi - P_{\delta_1})(\xi - P_{\gamma_2})(\xi - P_{\delta_2})}},$$

$$(113) \qquad M_3(\xi) = c_3 \frac{\xi - a}{\sqrt{(\xi - P_{\gamma_1})(\xi - P_{\delta_1})(\xi - P_{\gamma_2})(\xi - P_{\delta_2})}},$$

where in each instance, the quantity $\sqrt{(\xi - P_{\gamma_1})(\xi - P_{\delta_1})(\xi - P_{\gamma_2})(\xi - P_{\delta_2})}$ is taken to be analytic in $\mathbb{C} \setminus (\Gamma_1^+ \cup \Gamma_2^+)$, and the constants c_1 and c_2 are taken to be

$$(114) \qquad c_1 = \frac{1}{2a}\sqrt{(a - P_{\gamma_1})(a - P_{\delta_1})(a - P_{\gamma_2})(a - P_{\delta_2})},$$

$$(115) \qquad c_2 = \frac{-1}{2a}\sqrt{(-a - P_{\gamma_1})(-a - P_{\delta_1})(-a - P_{\gamma_2})(-a - P_{\delta_2})}.$$

We will require the following 3×3 matrix $P(z)$:

$$(116) \qquad P(z) = \begin{pmatrix} M_1(r_0(z)) & M_1(r_1(z)) & M_1(r_2(z)) \\ M_3(r_0(z)) & M_3(r_1(z)) & M_3(r_2(z)) \\ M_2(r_0(z)) & M_2(r_1(z)) & M_2(r_2(z)) \end{pmatrix}.$$

For future reference, the matrix $P(z)$ solves the following Riemann–Hilbert problem.

Riemann-Hilbert Problem 7.1.   (a) *P is analytic on $\mathbb{C} \setminus (I_1 \cup I_2)$.*
     (b) *The boundary values of P satisfy*

$$(117) \qquad P_+(x) = P_-(x)\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \textit{for } z \in I_1,$$

$$(118) \qquad P_+(x) = P_-(x)\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad \textit{for } z \in I_2,$$

(119)

     (c) *As $z \to \infty$, we have*

$$(120) \qquad P(z) = \left(I + O\left(\frac{1}{z}\right) \right).$$

Next we draw 4 small circles, centered at each of the endpoints $\gamma_1, \delta_1, \gamma_2, \delta_2$, and decompose the plane again, as shown in Figure 2.

Figure 2. Four disks, centered at each of the endpoints of the intervals I_1 and I_2. We decompose the plane into five regions: the four disks and the one exterior region.

We will now define a matrix D_{approx} which will be a uniformly valid approximation to the matrix D. The matrix D_{approx} will be defined separately in each of the 5 regions as shown in Figure 2. We set

$$(121) \qquad D_{approx}(z) = P(z), \quad \text{for } z \text{ outside the four disks.}$$

For z in the interior of each of the four disks, D_{approx} is defined in a (by now) standard way using the solutions of the Airy equation. Rather than presenting the details for each of the disks, we will focus only

19

on the disk centered at δ_1, as the construction for each of the other disks may be carried out by very similar calculations. Define

$$(122) \qquad S(z) = \left[\frac{3}{4} \int_{\delta_1}^{z} (2f_1(x) + f_2(x) - V_1'(x)) \, dx \right]^{2/3},$$

taken to be analytic in a neighborhood of $z = \delta_1$, and with $S'(\delta_1) > 0$ (which is possible because of conditions E1 and E2). It turns out that with this definition, one may take an appropriate branch so that $S(z)^{3/2} = -\frac{3i}{4}\theta_1(z)$ for z in the upper lens shaped region above I_1, and $S(z)^{3/2} = \frac{3i}{4}\theta_1(z)$ for z in the lower lens shaped region below I_1.

Next, define the matrix $\Phi(S)$ as follows (cf. [3, (7.13)]).

$$(123) \qquad \Phi(s) = \begin{cases} \begin{pmatrix} y_0(s) & -y_2(s) & 0 \\ y_0'(s) & -y_2'(s) & 0 \\ 0 & 0 & 1 \end{pmatrix}, & \text{for } 0 < \arg s < \frac{2\pi}{3}, \\[2em] \begin{pmatrix} -y_1(s) & -y_2(s) & 0 \\ -y_1'(s) & -y_2'(s) & 0 \\ 0 & 0 & 1 \end{pmatrix}, & \text{for } \frac{2\pi}{3} < \arg s < \pi, \\[2em] \begin{pmatrix} -y_2(s) & y_1(s) & 0 \\ -y_2'(s) & y_1'(s) & 0 \\ 0 & 0 & 1 \end{pmatrix}, & \text{for } -\pi < \arg s < -\frac{2\pi}{3}, \\[2em] \begin{pmatrix} y_0(s) & y_1(s) & 0 \\ y_0'(s) & y_1'(s) & 0 \\ 0 & 0 & 1 \end{pmatrix}, & \text{for } -\frac{2\pi}{3} < \arg s < 0. \end{cases}$$

where $y_0, y_1,$ and y_2 are the following solutions of the Airy equation (recall that $\mathrm{Ai}(x)$ is the unique solution to Airy's equation $y'' = xy$ which is exponentially decaying for $x \to +\infty$):

$$(124) \qquad y_0(s) = \mathrm{Ai}(s), \quad y_1(s) = \omega \mathrm{Ai}(\omega s), \quad y_2(s) = \omega^2 \mathrm{Ai}(\omega^2 s)$$

(here $\omega = e^{\frac{2\pi i}{3}}$).

And then define $E_n(z)$ for z in the disk centered at δ_1 (again following [3]) as follows:

$$(125) \qquad E_n(z) = \sqrt{\pi} P(z) \begin{pmatrix} 1 & -1 & 0 \\ -i & -i & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} n^{1/6} S(z)^{1/4} & 0 & 0 \\ 0 & n^{-1/6} S(z)^{-1/4} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Finally then, we may define $D_{approx}(z)$ for z in the disk centered at δ_1:

$$(126) \qquad D_{approx}(z) = E_n(z) \Phi\left(n^{2/3} S(z) \right) \mathrm{diag}\left(e^{-\frac{n_1}{2}\left(g_1^+ - g_1^- \right)}, \ e^{\frac{n_1}{2}\left(g_1^+ - g_1^- \right)}, \ 1 \right).$$

There is a similar definition of D_{approx} in each of the other 3 disks shown in Figure 2, which are entirely similar to the construction presented here [3]. We will omit these definitions here.

The matrix D_{approx} is a piecewise analytic function in the entire plane, and it is a global approximation to the matrix $D(z)$ defined in Figure 1. The proof of this follows a by now straightforward procedure, which has been described in [3, (8.1)-(8.6)]. Indeed, setting

$$(127) \qquad R(z) = D(z) D_{approx}^{-1}(z),$$

one verifies that $R(z)$ satisfies a Riemann–Hilbert problem of exactly the same type as the analogous function $R(z)$ defined in [3, (8.1)]: jumps across the boundaries of the disks are $I + \mathcal{O}(n^{-1})$, and jumps across all other contours are $I + \mathcal{O}(e^{-cn})$, and in addition the actual contours chosen may be deformed slightly (as in the text between (8.5) and (8.6) of [3]). Therefore one may conclude that

$$(128) \qquad R(z) = I + \mathcal{O}\left(\frac{1}{n(|z| + 1)} \right) \text{ as } n \to \infty,$$

uniformly for $z \in \mathbb{C}$.

8. Asymptotics for eigenvalue statistics of random matrices with source

In this short section we will describe some of the results that follow immediately from the results of the previous section. The proofs of these results follow by considerations entirely similar to those carried out in the Gaussian case by Bleher and Kuijlaars [3]: one expresses the kernel $K_n(x, y)$ in terms of the explicit transformations leading to $R(z)$, whose asymptotic expansion for $n \to \infty$ is under control.

Theorem 8.1. *Suppose that the conditions described as the "ideal situation", items A-D, described in Section 2 are satisfied, along with the conditions described in E1 and E2 of Section 3. Suppose in addition that the sets I_1 and I_2 are each single intervals, with $I_1 = (\gamma_1, \delta_1)$ and $I_2 = (\gamma_2, \delta_2)$. Then the following results hold true.*

- *The mean density of states, $\rho_n(x) := K_n(x, x)$ converges, as $n \to \infty$, to a limiting mean density, supported on $I_1 \cup I_2$, with*

$$(129) \qquad \lim_{n \to \infty} \rho_n(x) = -\frac{1}{2\pi i} \frac{d}{dx} \left(g_j^+(x) - g_j^-(x) \right), \quad \text{for } x \in I_j, \quad j = 1, 2.$$

- *Bulk universality holds true: for every $x_0 \in (\gamma_1, \delta_1) \cup (\gamma_2, \delta_2)$, and every $u, v \in \mathbb{R}$, we have*

$$(130) \qquad \lim_{n \to \infty} \frac{1}{n\rho(x_0)} K_n \left(x_0 + \frac{u}{n\rho(x_0)}, x_0 + \frac{v}{n\rho(x_0)} \right) = \frac{\sin \pi(u - v)}{\pi(u - v)}.$$

- *Edge universality also holds true, at each endpoint of $I_1 \cup I_2$: There are constants $c_{\gamma_1}, c_{\delta_1}, c_{\gamma_2}$, and c_{δ_2} so that for $j = 1, 2$, the following statements hold true: for every $u, v \in \mathbb{R}$, we have*

$$(131) \qquad \lim_{n \to \infty} \frac{1}{(c_{\gamma_j} n)^{2/3}} K_n \left(\gamma_j + \frac{u}{(c_{\gamma_j} n)^{2/3}}, \gamma_j + \frac{v}{(c_{\gamma_j} n)^{2/3}} \right) = \frac{Ai(u)Ai'(v) - Ai(v)Ai'(u)}{u - v},$$

$$(132) \qquad \lim_{n \to \infty} \frac{1}{(c_{\delta_j} n)^{2/3}} K_n \left(\delta_j - \frac{u}{(c_{\delta_j} n)^{2/3}}, \delta_j - \frac{v}{(c_{\delta_j} n)^{2/3}} \right) = \frac{Ai(u)Ai'(v) - Ai(v)Ai'(u)}{u - v},$$

Remark: As proven in Section 6, all of the conditions A-D along with E1 and E2 hold true for all a sufficiently large, for the quartic case $V(x) = \frac{1}{4} x^4$, with n even and $n_1 = n_2 = n/2$. Thus bulk and edge universality have been extended beyond the Gaussian case for random matrices with source.

It is natural to expect that the same results should hold true for all a sufficiently large, under the more general assumption that the external field $V(x)$ is convex, real analytic, and with sufficient growth for $|x| \to \infty$. While it seems a rather daunting task to carry out the requisite asymptotic analysis for $a \to \infty$ to prove that a suitable curve exists (note that it need not be algebraic), there is a possibility to combine the analysis described herein with an analysis of the coupled variational problem described following formula (1.15) in [3].

References

[1] A. Aptekarev, P. Bleher, and A. Kuijlaars, Large n limit of Gaussian random matrices with external source, part II. *Comm. Math. Phys.* **259** (2005), 367-389.

[2] P. Bleher and A. Kuijlaars, Random matrices with external source and multiple orthogonal polynomials. *Int. Math. Research Notices* **2004**, no 3 (2004), 109 - 129.

[3] P. Bleher and A. Kuijlaars, Large n limit of Gaussian random matrices with external source, part I. *Commun. Math. Phys.* **252** (2004), 43-76.

[4] P. Bleher and A. Kuijlaars Large n limit of Gaussian random matrices with external source, part III: double scaling limit. *to appear in Comm. Math. Phys.*, (math.ph/0602064).

[5] E. Breézin and S. Hikami. Universal singularity at the closure of a gap in a random matrix theory. *Phys. Rev. E* **57** (1998), 4140-4149.

[6] E. Beézin and S. Hikami. Level spacing of random matrices in an external source. *Phys. Rev. E* **58** (1998), 7176-7185.

[7] P. Deift, T. Kriecherbauer, K. T.-R. McLaughlin, S. Venakides, and X. Zhou, "Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory", *Comm. Pure Appl. Math.*, **52**, 1335–1425, 1999.

[8] P. Deift, T. Kriecherbauer, K. T.-R. McLaughlin, S. Venakides, and X. Zhou, "Strong asymptotics of orthogonal polynomials with respect to exponential weights", *Comm. Pure Appl. Math.*, **52**, 1491–1552, 1999.

[9] A. Kuijlaars, W. van Assche, and F. Wielonsky, Quadratic Hermite-Padé approximation to the exponential function: a Riemann-Hilbert approach, preprint math.CA/0302357.

[10] M. L. Mehta, *Random Matrices*, Second edition, Academic Press, Boston, 1991.

[11] L. Pastur, The spectrum of random matrices (Russian), *Teoret. Mat. Fiz.* **10** (1972), 102-112.

[12] G. Szegő, *Orthogonal Polynomials*, (Fourth Edition), American Math. Society Colloquium Publications, Volume 23, Providence, 1975.

[13] P. Zinn-Justin, Random Hermitian matrices in an external field. *Nucl. Phys. B* **497** (1997), 725-732.

[14] P. Zinn-Justin, Universality of correlation functions of hermitian random matrices in an external field, *Commun. Math. Phys.* **194** (1998), 631, cond-mat/9705044.

K. T.-R. McLaughlin: Department of Mathematics, University of Arizona, 617 N Santa Rita Ave, Tucson, AZ 85721 Email address: MCL@MATH.ARIZONA.EDU